



# GWAS Catalog access with gwasrapid

## Introduction

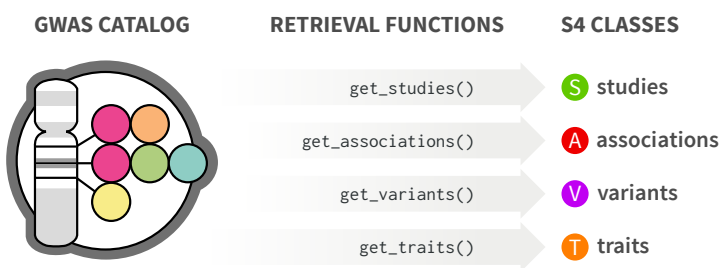
The **GWAS Catalog** is a service provided by the EMBL-EBI and NHGRI that offers a manually curated and freely available database of published genome-wide association studies (GWAS).

The GWAS Catalog data provided by the **RESTful API** is organized around four core entities:

- **studies**
- **associations**
- **variants**
- **traits**

## Get GWAS Catalog Entities

**gwasrapid** facilitates the access to the Catalog via the RESTful API, allowing you to programmatically retrieve data directly into R. Each of the four entities is mapped to an S4 object of a class of the same name.



Search by	Example	S	A	V	T
study_id	"GCST00858"	●	●	●	●
association_id	24300113	●	●	●	●
variant_id	"rs12752552"	●	●	●	●
efo_id	"EFO_0005543"	●	●	●	●
pubmed_id	"21626137"	●	●	●	●
user_requested	TRUE	●	●	●	●
full_pvalue_set	FALSE	●	●	●	●
efo_uri	"http://www.ebi.ac.uk/efo/EFO_0004761"	●	●	●	●
genomic_range	list(chromosome = "22", start = 1L, end = 15473564L)	●	●	●	●
gene_name	"BRCA1"	●	●	●	●
efo_trait	"lung adenocarcinoma"	●	●	●	●
reported_trait	"Breast cancer"	●	●	●	●

## S4 Representation of GWAS Catalog Entities

### S4 class studies

The **studies** object consists of eight slots, each a table (tibble). Each study is an observation (row) in the studies table — main table. All tables have the column study\_id as primary key.

For details about the studies S4 class: `class?studies`.

**studies**

- study\_id
- reported\_trait
- initial\_sample\_size
- replication\_sample\_size
- gxe
- gxg
- snp\_count
- qualifier
- imputed
- pooled
- study\_design\_comment
- full\_pvalue\_set
- user\_requested

**genotyping\_techs**

- study\_id
- genotyping technology

**platforms**

- study\_id
- manufacturer

**ancestries**

- study\_id
- ancestry\_id
- type
- number\_of\_individuals

**ancestral\_groups**

- study\_id
- ancestry\_id
- ancestral\_group

**countries\_of\_recruitment**

- study\_id
- ancestry\_id
- country\_name
- major\_area
- region

**countries\_of\_origin**

- study\_id
- ancestry\_id
- country\_name
- major\_area
- region

**publications**

- study\_id
- pubmed\_id
- publication\_date
- publication
- title
- author\_fullname
- author\_orcid

### S4 class associations

The **associations** object consists of six slots, each a table (tibble). Each study is an observation (row) in the associations table — main table. All tables have the column association\_id as primary key.

For details about the associations S4 class: `class?associations`.

**associations**

- association\_id
- pvalue
- pvalue\_description
- pvalue\_mantissa
- pvalue\_exponent
- multiple\_snp\_haplotype
- snp\_interaction
- standard\_error
- range
- or\_per\_copy\_number
- beta\_number
- beta\_unit
- beta\_direction
- beta\_description
- last\_mapping\_date
- last\_update\_date

**loci**

- association\_id
- locus\_id
- haplotype\_snp\_count
- description

**risk\_alleles**

- association\_id
- locus\_id
- variant\_id
- risk\_allele
- risk\_frequency
- genome\_wide
- limited\_list

**genes**

- association\_id
- locus\_id
- gene\_name

**ensembl\_ids**

- association\_id
- locus\_id
- gene\_name
- ensembl\_id

**entrez\_ids**

- association\_id
- locus\_id
- gene\_name
- entrez\_id

### S4 class variants

The **variants** object consists of four slots, each a table (tibble). Each variant is an observation (row) in the variants table — main table. All tables have the column variant\_id as primary key.

For details about the variants S4 class: `class?variants`.

**variants**

- variant\_id
- merged
- functional\_class
- chromosome\_name
- chromosome\_position
- chromosome\_region
- last\_update\_date

**genomic\_contexts**

- variant\_id
- gene\_name
- chromosome\_name
- chromosome\_position
- distance
- is\_closest\_gene
- is\_intergenic
- is\_upstream
- is\_downstream
- source
- mapping\_method

**ensembl\_ids**

- variant\_id
- gene\_name
- ensembl\_id

**entrez\_ids**

- variant\_id
- gene\_name
- entrez\_id

### S4 class traits

The **traits** object consists of one slot only, a table (tibble) of GWAS Catalog EFO traits. Each EFO trait is an observation (row) in the traits table — main table.

For details about the traits S4 class: `class?traits`.

**traits**

- efo\_id
- trait
- uri

## Manipulate Cases

Get a **studies** object **s** of two GWAS studies:

```
s <- get_studies(study_id = c('a', 'b'))
```

Subset object **s** by either identifier or position using ``[``:

```
s['a'] # Subset by identifier
```

```
s[1] # Subset by position
```

Combine two studies' objects:

```
union(s1,s2)
```

(Discards duplicates)

```
bind(s1,s2)
```

(Keeps duplicates)